Search for $t\bar{t}$ resonance in ATLAS

William BARBE
Laboratoire de Physique de Clermont

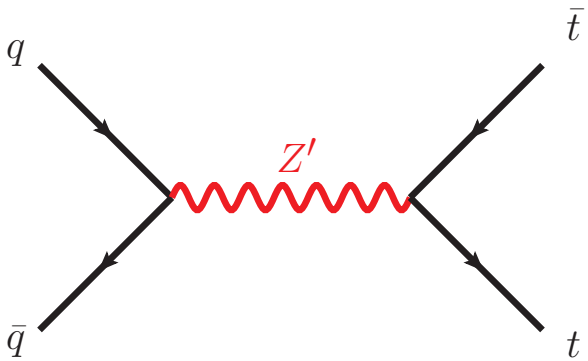
Thursday, April 25^{th} 2019

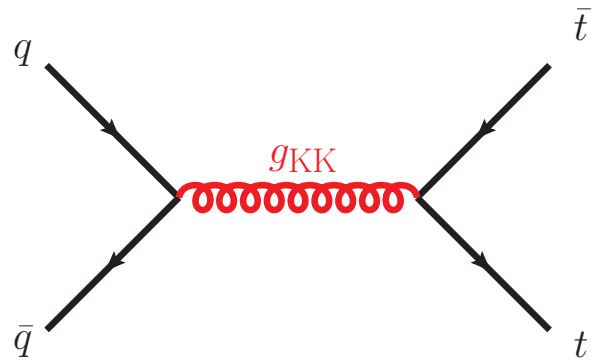
1. Introduction
2. Results with partial run 2 with 36.1 fb^{-1} data
3. Functional decomposition
4. Strategy to validate FD
5. Summary

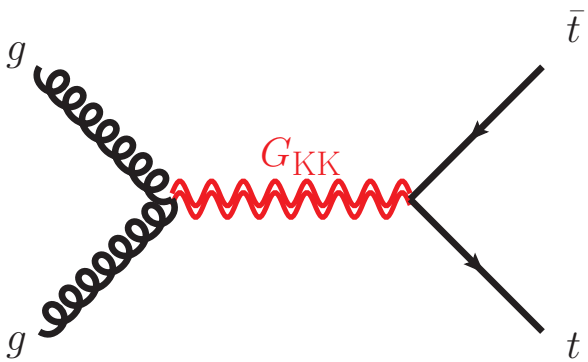
$t\bar{t}$ l+jets analysis in a nutshell

Search for new resonance decaying into a top quark pair

- ⋆ Search for resonances in the $t\bar{t}$ mass spectrum predicted by BSM theories (Z', KK gluon, KK graviton ...)
- ⋆ 1 lepton top-antitop final state $t\bar{t} \rightarrow WbWb \rightarrow l\nu bqq'b$
- ⋆ Signature with high p_T lepton, large MET and hadronic jets





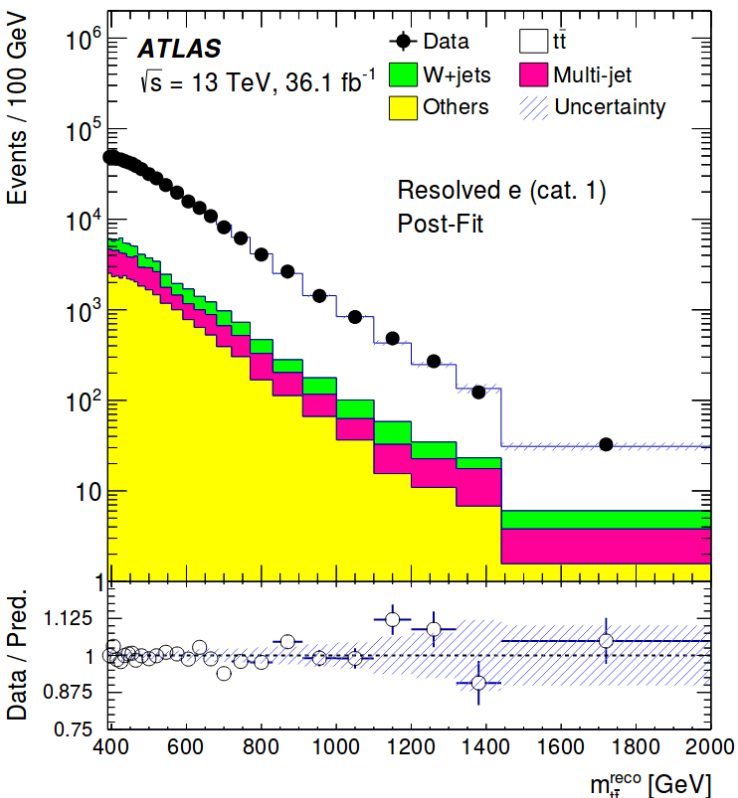


Analysis presented in the last Top LHC France

I will just give a summary of the results using partial run 2 data (36.1 fb^{-1}) and then show the methods that are being investigated for bkg estimation using full run 2 data

Results with partial run 2 data (36.1 fb^{-1})

Type	Yields			
	Boosted e	Boosted μ	Resolved e	Resolved μ
$t\bar{t}$	28 500±500	26 000±400	231 100±1900	225 300±1700
W+jets	2200±240	2200±180	9400±1100	10 300±800
Multi-jet	2000±400	780±200	8200±1400	7400±1400
Others	2880±230	2420±180	13 000±600	12 000±500
Total	35 600±500	31 300±300	262 200±1200	254 600±1100
Data	35612	31188	261554	254277



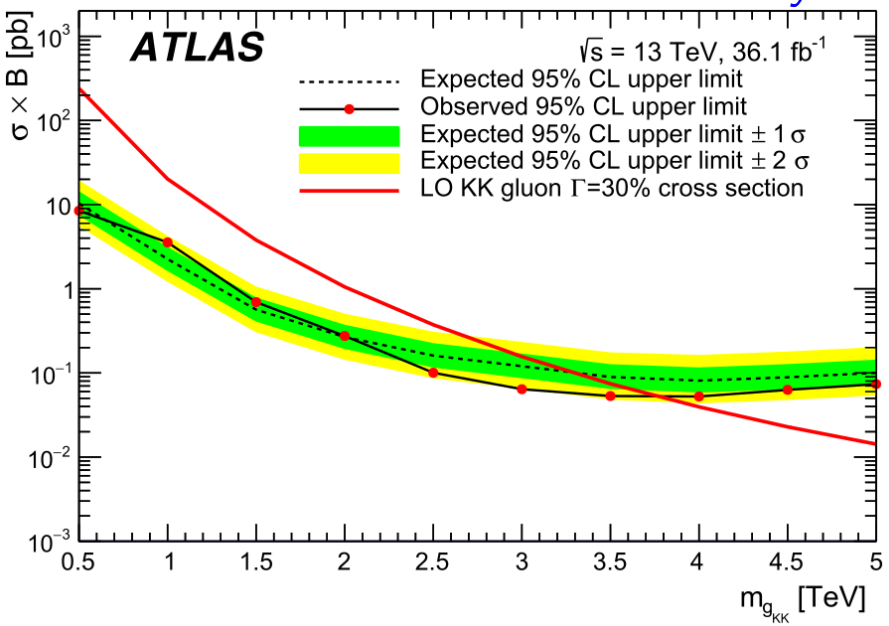
- ⋆ Search for excesses in the top-antitop mass spectrum
  - good agreement in all the 12 signal regions
  - exclusion limits set on benchmark models

https://arxiv.org/abs/1804.10823
Eur. Phys. J. C 78 (2018) 565

Results with partial run 2 data (36.1 fb^{-1})

- ⋆ Limits are set on benchmark model productions cross-sections :
  - → Z', KK gluon (g_{KK}), KK graviton

https://arxiv.org/abs/1804.10823
Eur. Phys. J. C 78 (2018) 565



Summary of 95 % Confidence Level mass exclusion ranges on benchmark models		
Model	Observed excluded mass [TeV]	Expected excluded mass [TeV]
Z'_{TC2} (1% width)	< 3.0	< 2.6
$Z'_{\mathrm{DM,ax}}$	< 1.2	< 1.4
$Z'_{\mathrm{DM,vec}}$	< 1.4	< 1.6
G_{KK}	[0.45, 0.65]	[0.45, 0.65]
g_{KK} (15% width)	< 3.8	< 3.5
g_{KK} (30% width)	< 3.7	< 3.2

From 36.1 to 150 fb^{-1}

- ⋆ 36.1 fb^{-1} analysis : backgrounds mostly from MC samples
- ⋆ $W+$ jets and multijet contributions were estimated from data :
  - $W+$ jets → scale factors derived from data, applied to correct the normalization given by MC simulation
  - multijet → estimated using the matrix method (trickier and trickier when the trigger isolation get close to the analysis one)
- ⋆ O(100) systematics, 12 channels with O(20) bins and large statistic
  - ⇒ profiling is very challenging
  - ⇒ more than 6 months to tune the fit

- ⋆ For full run 2 data : try data-driven bkg estimate → Functional Decomposition (FD)
  - avoid all the above issues
  - using (almost) only the data (MC needed only for the signal and tests)

Functional decomposition (FD)

- ⋆ Method to fit falling smooth background
- ⋆ Decompose data into moments : use first few moments for bkg estimation
- ⋆ Higher moments used to estimate the resonants contributions
- ⋆ FD's paper : https://arxiv.org/pdf/1805.04536.pdf

Advantages

- ⋆ No fake estimate
- ⋆ no more need to spend months to tune the fit
- ⋆ Using (almost) only the data (MC needed only for the signal and to validate the method)
- ⋆ Can in principle represent any shape
- ⋆ Model full spectra

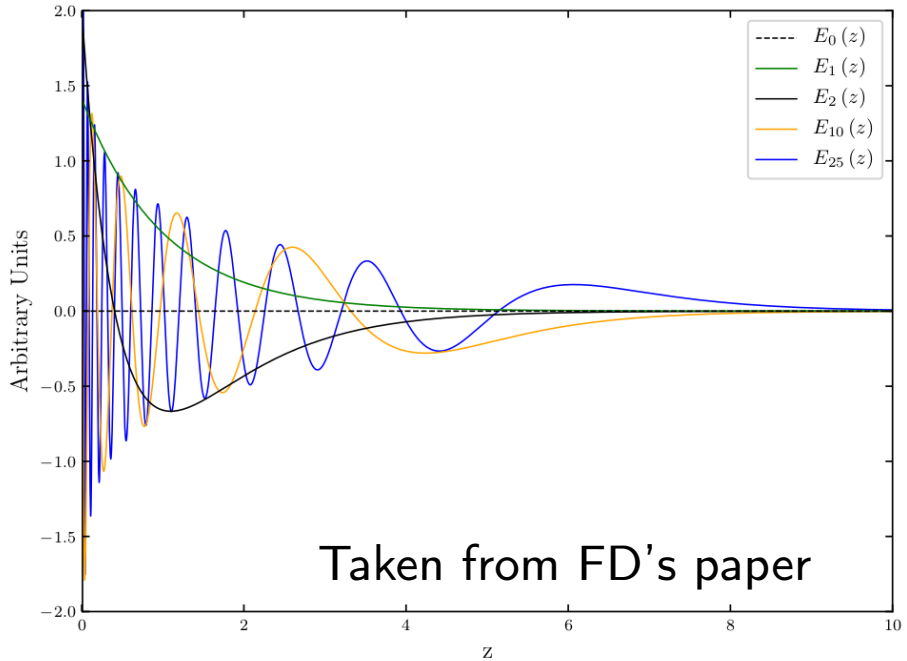
Basics of FD

- ⋆ Based on a set of complete, orthonormal functions
  - orthonormalize $F_n(z) = \sqrt{2}e^{-nz}$
  - Solution :

$$\phi_n = \sqrt{1 - \frac{1}{n^2}}$$

$$E_1(Z) = \sqrt{2}e^{-z}$$

$$E_{n+1}(z) = (4e^{-z} - \frac{2}{\phi_{2n}^2})\frac{E_n(z)}{\phi_{2n+1}} + \phi_{2n-1}\frac{E_{n-1}(z)}{\phi_{2n+1}}$$



- ⋆ Recursion relations → fast evaluation of $E_n(z)$

Coordinate transform

- We need to ensure that the tail is well modeled :
  - all orthonormal exponentials approach e^{-z} as $z \to \infty$
  - hyperparameters adjust the shape of the tail

- To do that, a coordinate transform is used :

$$z = \left(\frac{x - x_0}{\lambda} \right)^{\alpha}$$

  - x is the variable of interest ($m_{t\bar{t}}$)
  - z is the corresponding dimensionless variable
  - dataset $\{x_m\} \Leftrightarrow \{z_m\}$

- Hyperparameters :
  - x_0 : lower mass cut (offset)
  - λ : mass scale
  - α : dimensionless exponent

Choice of hyperparameters crucial for FD's efficiency

Hyperparameters optimization

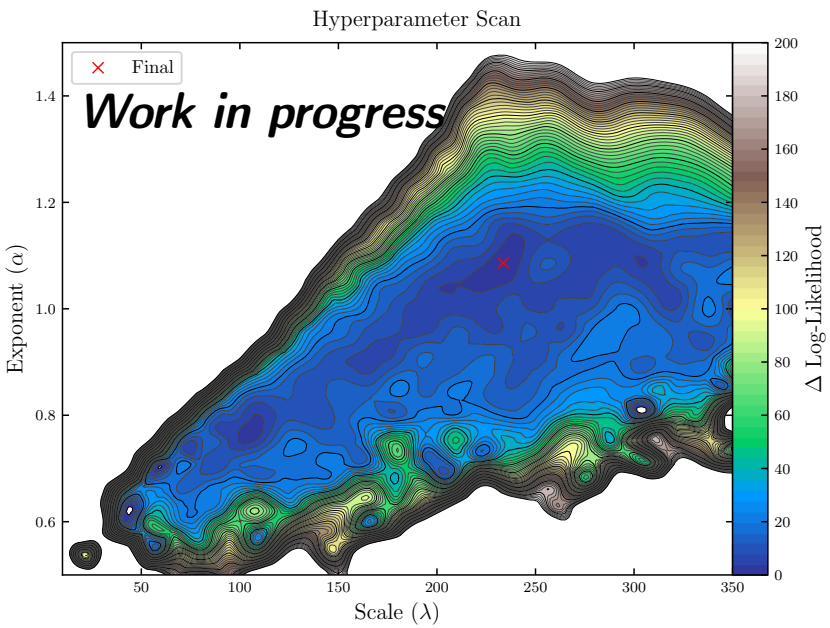
- ⋆ Selection of λ and α can greatly affect the number of terms $\mathcal{N}$ needed to model the background
- ⋆ Hyperparameters are chosen in order to minimize :

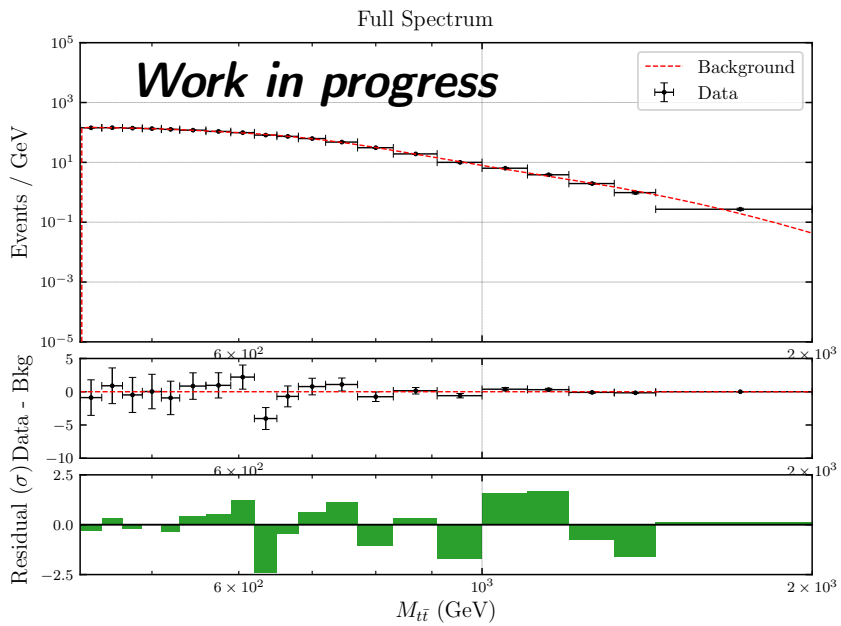
$$\mathcal{L} = \text{LogP}(\text{Data, Model}) + ln(\frac{M}{\mathcal{N}e})$$

  - LogP : represents the amount of information to encode the data given the model (ie the compatiblity of the data with the model)
  - Penalty term : amount of information to encode the model

Fit on $m_{t\bar{t}}$

- ⋆ Dataset $\{z_m\}$ of M unbinned datapoints : $\Omega(z) = \sum_{n=0}^{\mathcal{N}-1} f_n E_n(z)$
  - f_n : coefficients of the background distribution
  - $E_n(z)$: the orthonormal exponentials
- ⋆ Fit on $m_{t\bar{t}}$: pseudo-data made of $t\bar{t}$ and W+jets MC samples
- ⋆ FD searches for the best (λ, α) :
  - test various number of moment $\mathcal{N}$ for bkg modeling
  - the couple (λ, α) and $\mathcal{N}$ that give the minimal $\mathcal{L}$ are chosen





In all the presentation : FD is used on **pseudo-datas**

Possible to extract the syst. on the fit (cf slide 14)

Including signal contributions

- Dataset $\{z_m\}$ of M unbinned datapoints :

$$\Omega(z) = \sum_{n=0}^{\mathcal{N}-1} c_n E_n(z) + \sum_{m=0}^{N_s} s_m S_m(z)$$

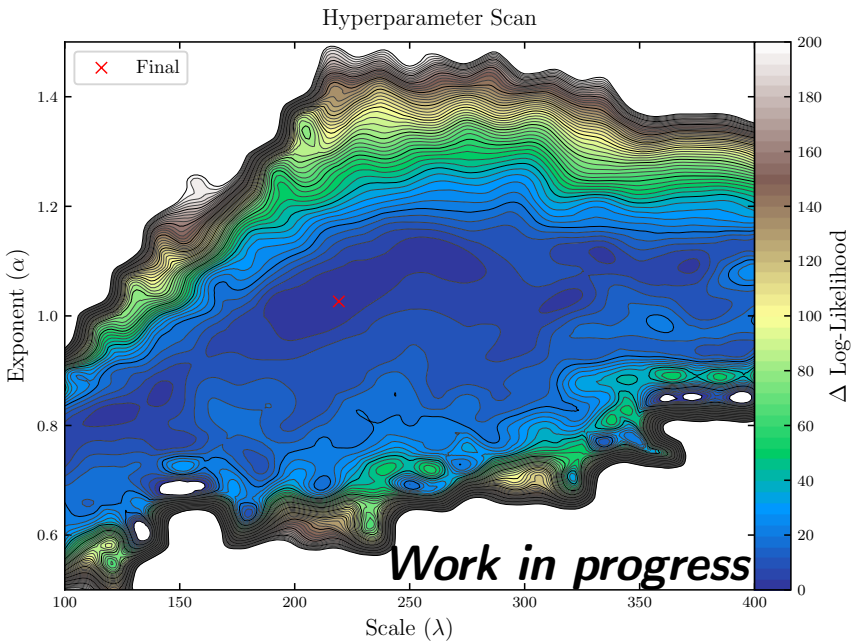
  - c_n : coefficients of the background distribution
  - $E_n(z)$: the orthonormal exponentials
  - s_m : signal normalization
  - $S_m(z)$: number of N_s resonant contributions

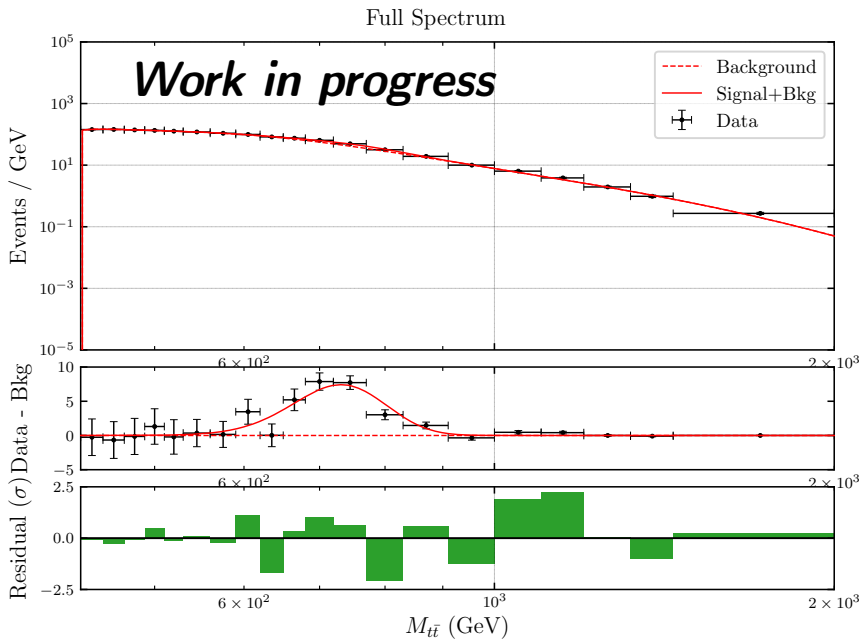
- First few $\mathcal{N}$ moments are enough to describe the bkg
  - $c_n = 0$ if $n \geq \mathcal{N}$

- Estimating c_n and s_m with the **method of the moments**
  - decompose the data into moments $\tilde{f}_n$
  - extract the signal contributions
  - bkg coefficients : $c_n = \tilde{f}_n - s_m \tilde{S}_{(m)n}$, $n < \mathcal{N}$

Fit on $m_{t\bar{t}} + Z'$

- ⋆ Fit on $m_{t\bar{t}}$:
  - same pseudo-data as before ($t\bar{t}$ and W+jets)
  - a Z' of 750 GeV has been injected in the pseudo-data
- ⋆ Knowing where the signal is, possible to fit it (here assuming a gaussian shape for the signal)
- ⋆ FD searches for the best (λ, α)



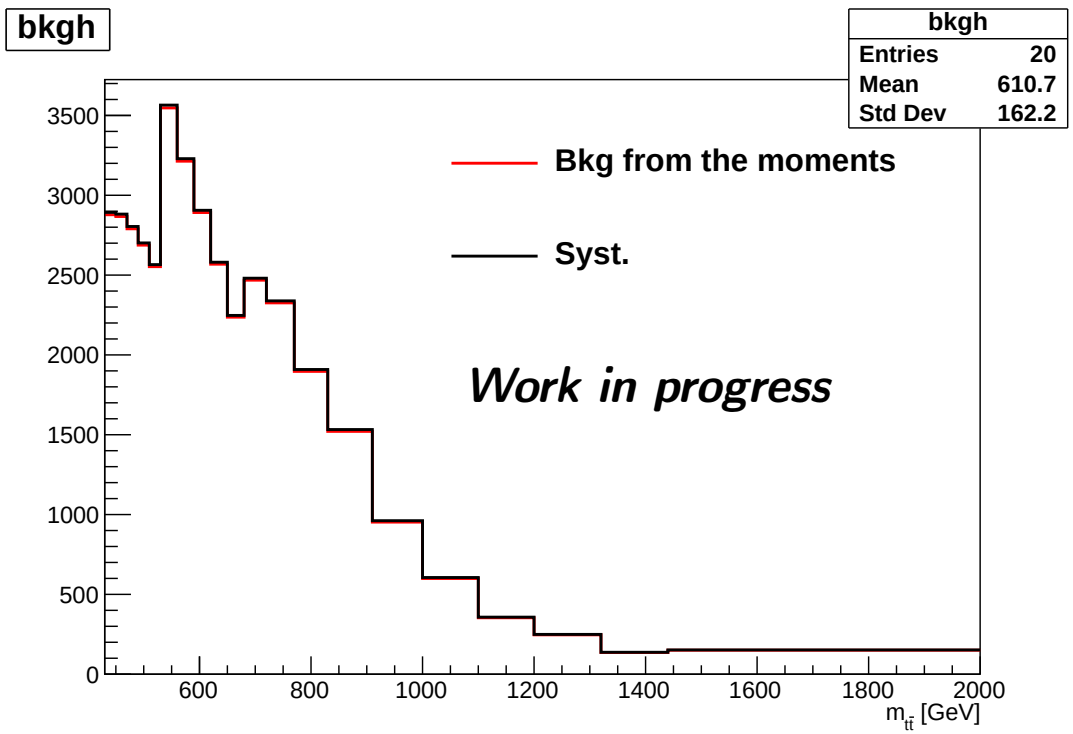


In all the presentation : FD is used on **pseudo-datas**

Possible to extract the syst. on the fit (cf slide 14)

Errors on the fit

- ⋆ Errors on the fit are accessible
  - → covariance matrix computed by FD
- ⋆ These errors can be used as systematics when running BH



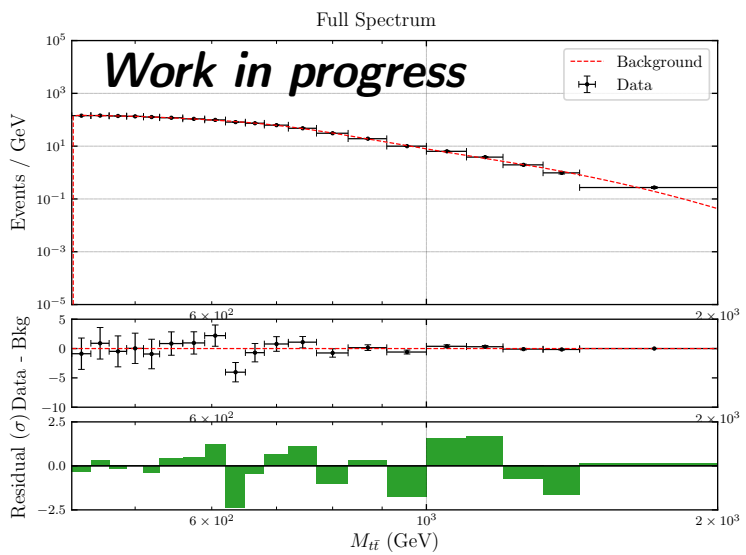
Strategy to validate FD

- ⋆ Check FD does not produce spurious signal :
  - generate many pseudo-data under B-only hypothesis → estimate bkg shape from FD
  - search for the most significant bump, and evaluate its significance
- ⋆ Check FD does not absorb the signal in the fit
  - inject various signal (mass, width, strength)
  - compare FD's performances with MC based analysis

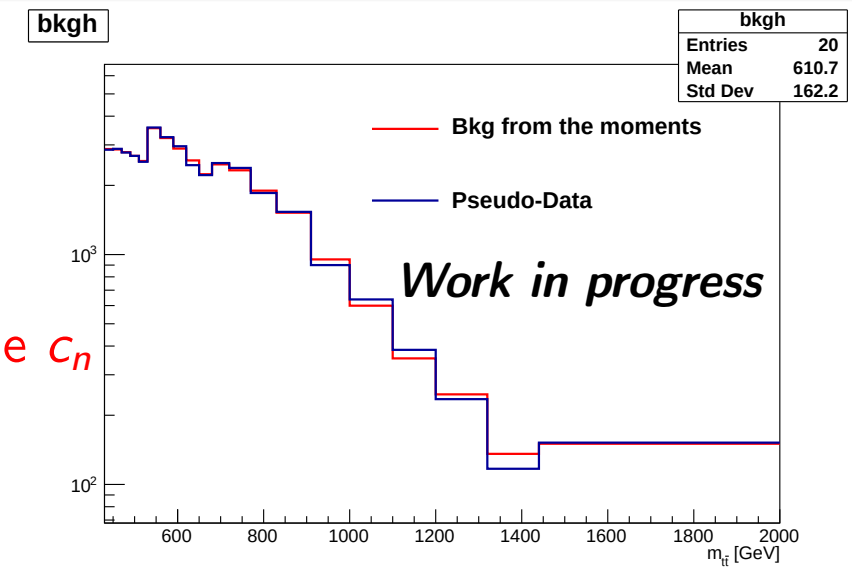
Strategy for the spurious signal study (1)

Get the bkg estimate from FD

- ⋆ Idea : FD + *BumpHunter*
- ⋆ First, run FD on pseudo-data
- ⋆ Get the moments c_n used by FD to model the bkg
- ⋆ Reconstruct the background estimate from the fit
- ⋆ Convert it in histogram



get the c_n

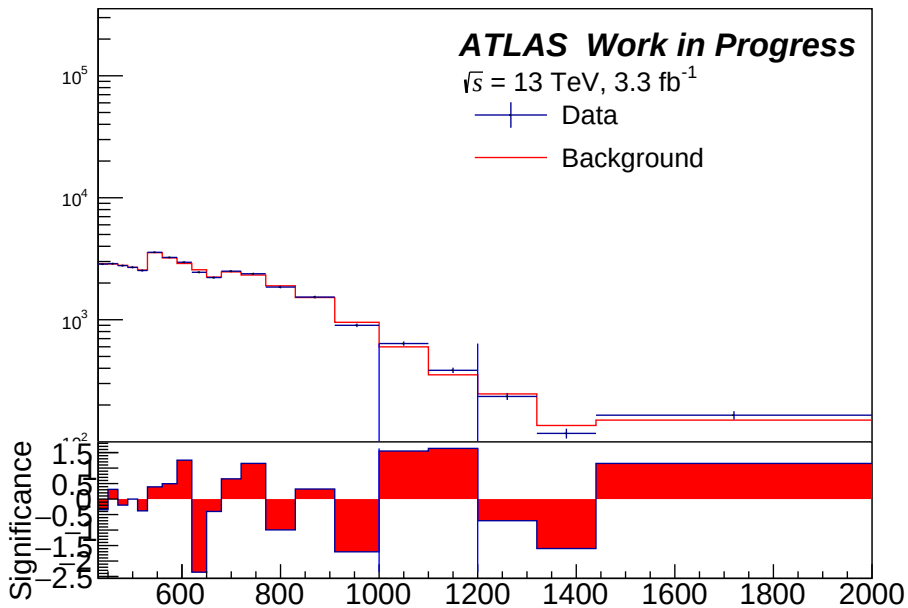


Strategy for the spurious signal study (2)

- ⋆ We have the bkg estimate and the pseudo-data
- ⋆ Run *BumpHunter* with the bkg estimate from FD (as we would do in the analysis)

BumpHunter returns the intervall with the most significant excess/deficit with a global p-value

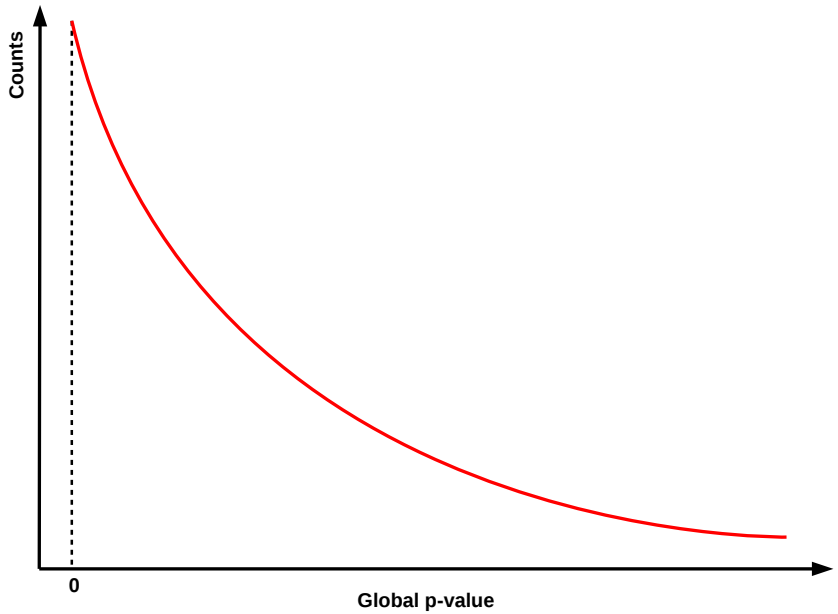
Repeat the procedure for several pseudo-data sets
⇒ distribution of global p-values



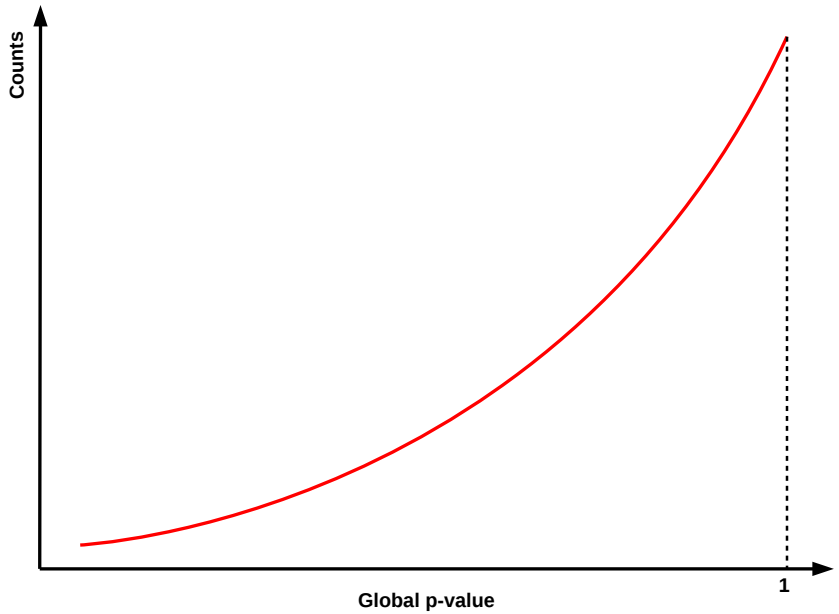
$$\text{p-val}^{local} = 0.013$$
$$\text{p-val}^{global} = 0.205$$

Strategy for the spurious signal study (3)

- ⋆ The shape of the distribution tell us if the fit is creating spurious signal
- ⋆ If the distribution is bias toward 0 → spurious signal ! !
- ⋆ If the distribution is bias toward 1 → FD is fitting the fluctuations (and potentially the signal ! !)



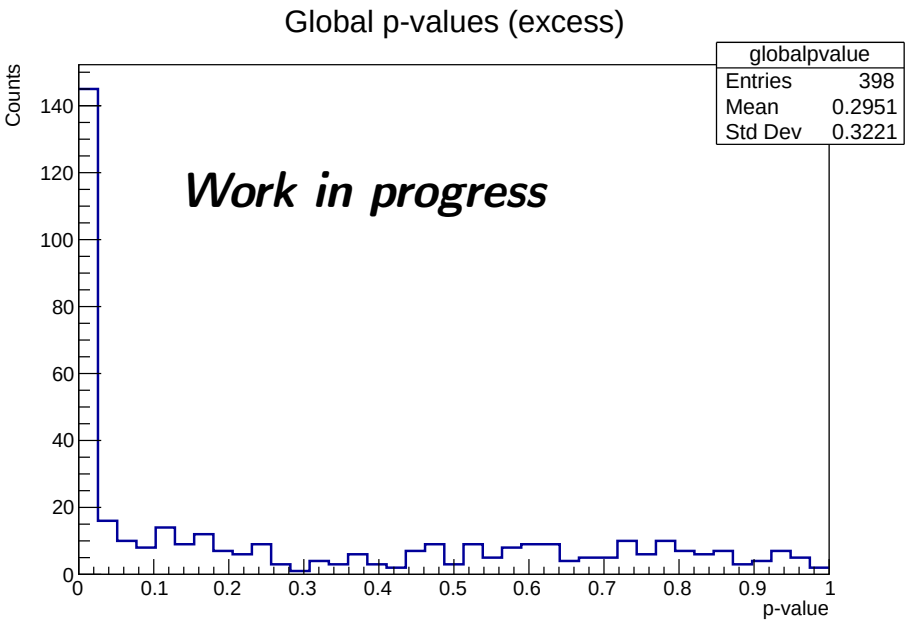
distributon bias toward 0
→ *BumpHunter* found a large discrepency for a large fraction of pseudo-data



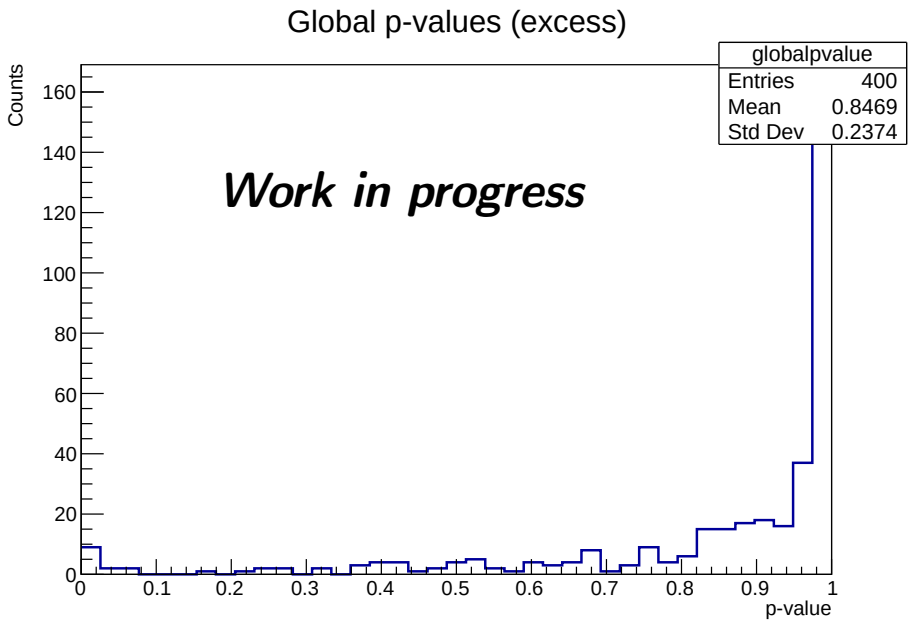
distribution bias toward 1
→ FD is probably using too much moments for bkg estimate

Distributions of global p-values

Using old FD



Using new FD



Old and New FD $\rightarrow$ 2 ways of defining the likelihood used for the hyperparameters optimization

- ⋆ Peak at 0 for the old FD :
  - BH see large discrepencies
  - the old FD creates spurious signal
- ⋆ Peak at 1 for the new FD :
  - the new FD is more able to adjust the data when there is no signal
  - but when there is a signal, fit it as bkg

Signal injection studies

- ⋆ Check FD does not absorb the signal in the fit
  - inject various signal (mass, width, strength)
  - compare FD's performances with MC based analysis

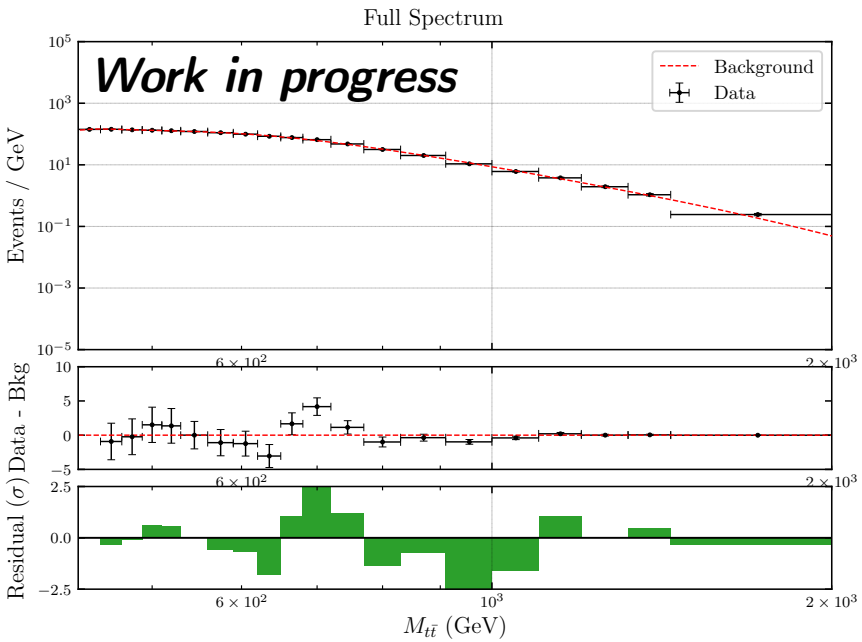
Compare limit sentitivity : MC vs FD

- ⋆ Compare exclusion upper limit MC vs FD
  1) pseudo-data from MC samples → "data"
  2) Bkg is either :
     - $t\bar{t}$ and W+jets MC samples
     - FD's bkg estimate

     ⇒ expected and observed limits for MC and FD
- ⋆ 2 scenarios for the "data" :
  - SM only
  - SM+Z' signal
- ⋆ Compare the 2 versions of FD (old and new)

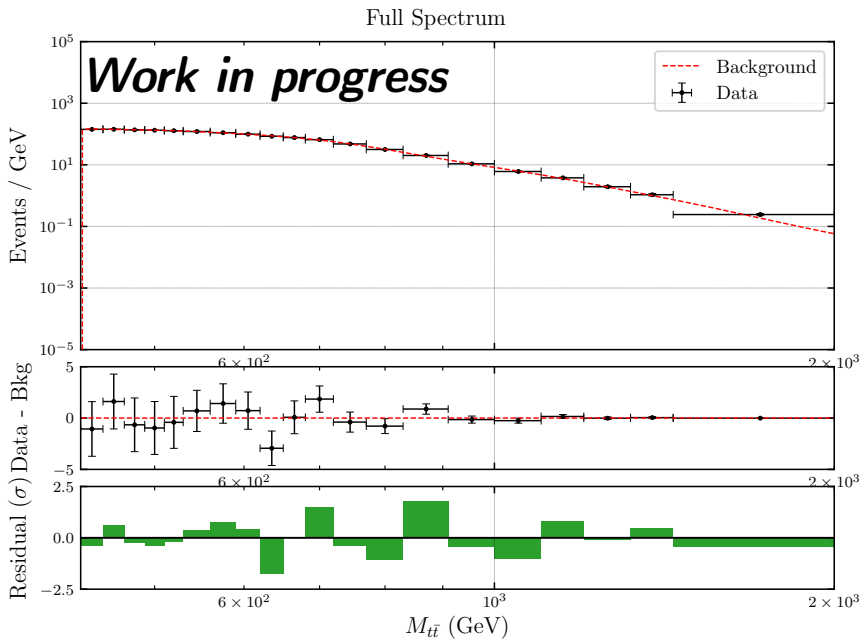
Old vs new FD : fit examples, resolved muon, btag category $3 + Z'$ of 750 GeV

Old FD



- ⋆ Excess around 700 GeV ?
- ⋆ Wave structure in the fit :
  - the signal is perturbing the fit

New FD



- ⋆ No excess in the fit
- ⋆ New FD is fitting the signal as bkg

Likelihood definition affect the capability of FD to see or not a signal

Some remarks on FD so far

- ⋆ Old FD is creating spurious signal
- ⋆ The new FD fit the signal as bkg :
  - more difficult to find a minimum (valley less clear in the hyperparameter scan)
  - new FD is more able to adjust the data (good to model the turn-on)
  - high risk to hide the signal
- ⋆ Now working on a better definition for the likelihood

Summary

- ⋆ Results with 36.1 fb^{-1} $\rightarrow$ no new physics discovered
  - results were used to set limits on benchmark models production cross-section
- ⋆ Classical ways for bkg modeling are becoming more and more difficult to use with the increase of statistics
- ⋆ Need to find new ways to model the bkg

Functional Decomposition

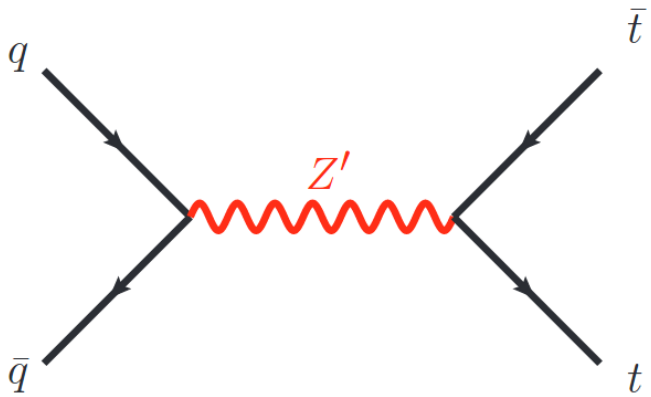
- ⋆ FD is a tool to fit data and search for new particles
- ⋆ FD has several pros :
  - use only the data
  - can in principle represent any shape
  - no need for MC (except for signal modelization)
- ⋆ Currently testing FD in $t\bar{t}$ l+jets :
  - spurious signals
  - signal injection tests

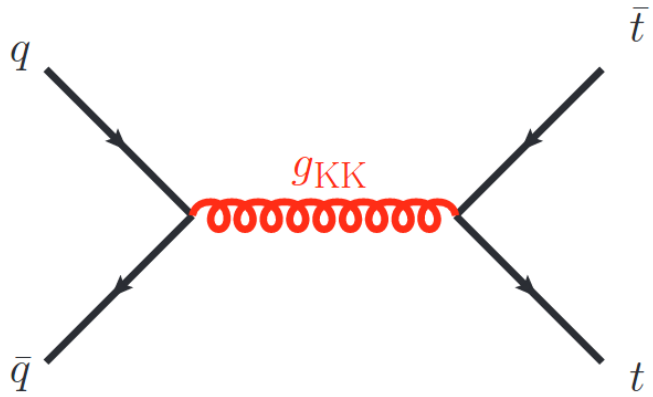
BACK-UP

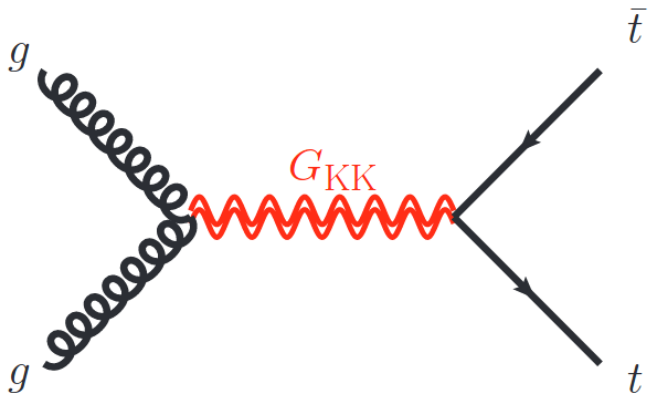
$t\bar{t}$ analysis

⋆ Search for resonances in the $t\bar{t}$ mass spectrum

→ Particles predicted by BSM theories (Z', KK gluon, KK graviton, 2HDM ...)

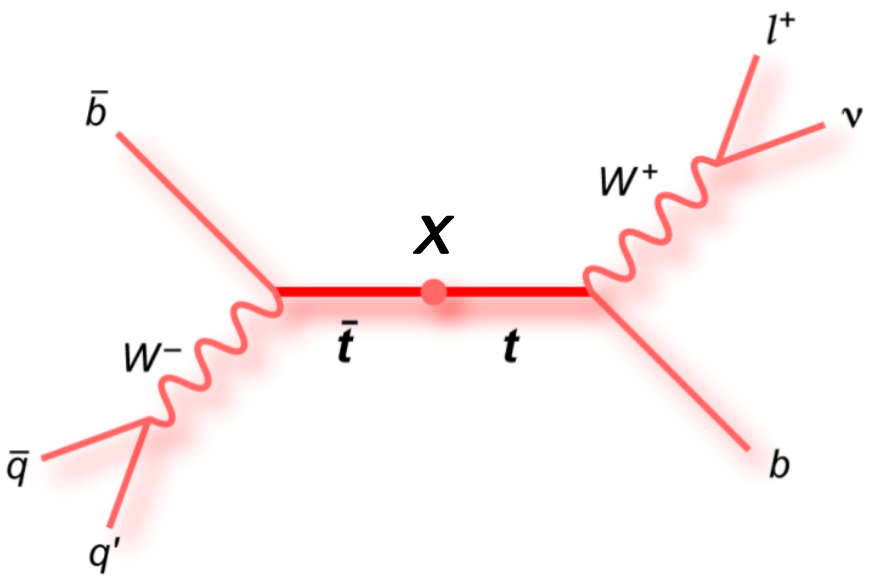


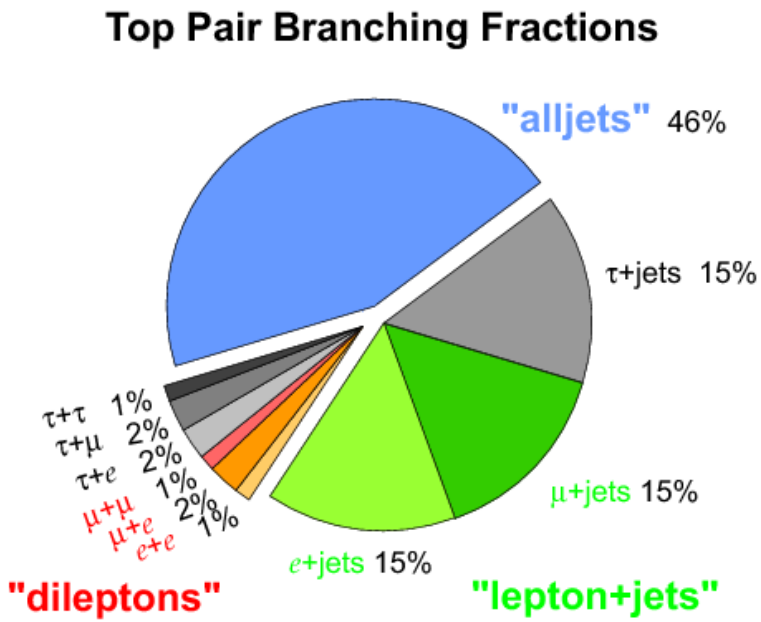




$t\bar{t}$ analysis

- Focus on semi-leptonic top-antitop final state $t\bar{t} \rightarrow WbWb \rightarrow l\nu bqq'b$
  - signature with high transverse momentum lepton, large MET and hadronic jets
- 36.1 fb^{-1} of data at 13 TeV (2015+2016)



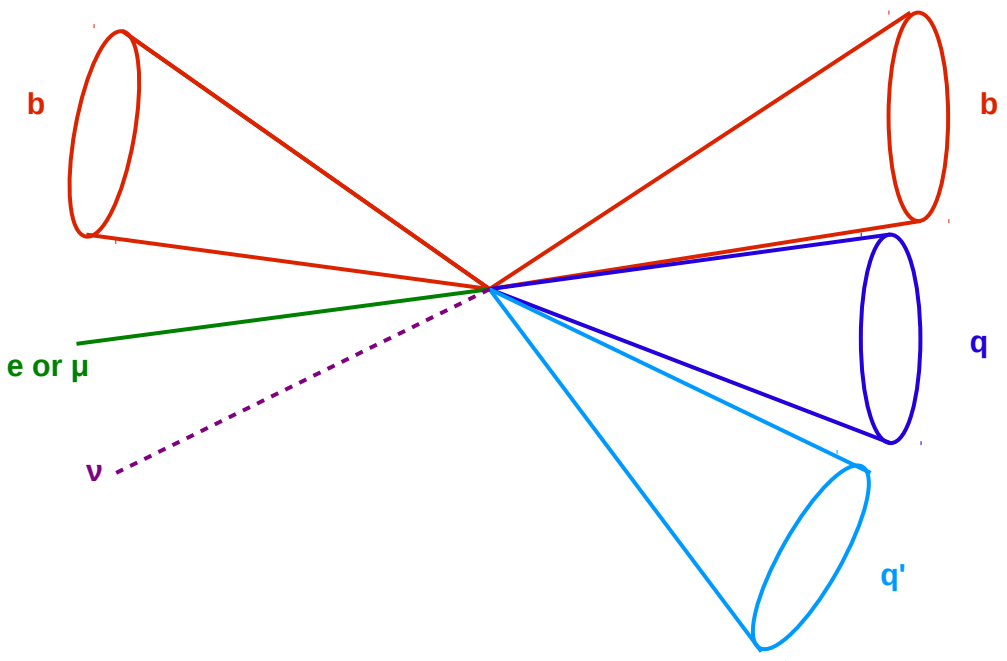


Event selection

- ⋆ Exactly one electron or muon
- ⋆ Missing transverse energy (MET)
- ⋆ At least 1 jet identified as a jet from a b quark (b-tagged)

Resolved selection

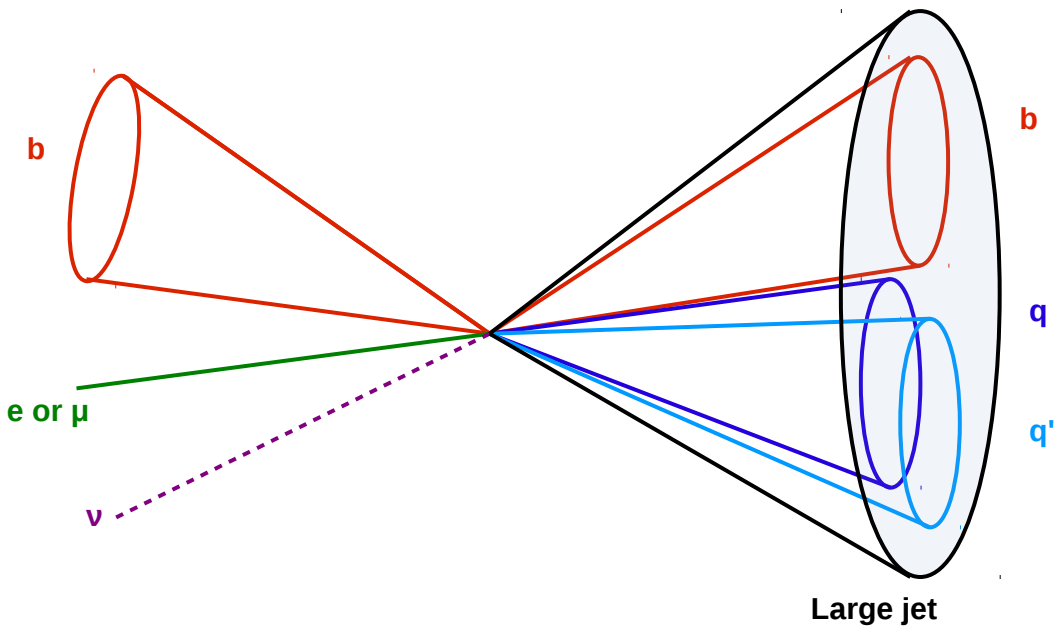
- ⋆ $\geq$ 4 small jets



Event selection

Boosted selection

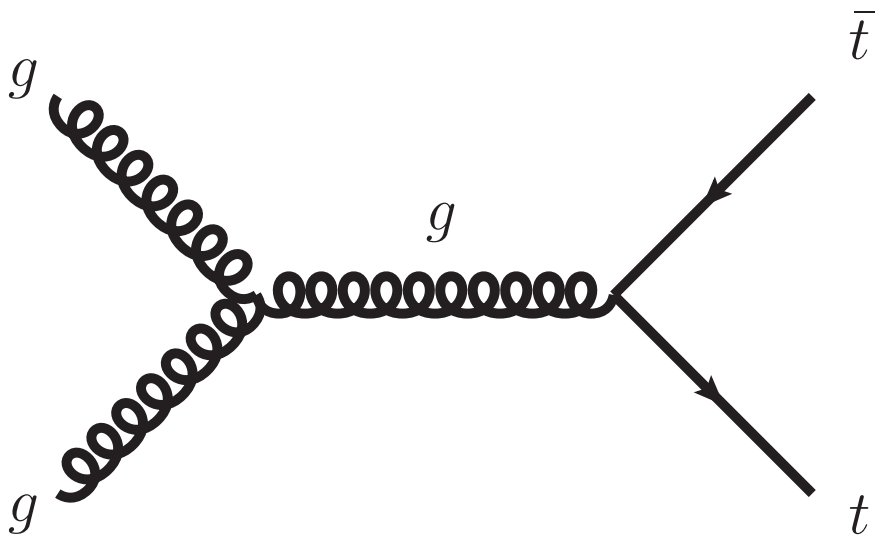
- ⋆ One large-R jet identified as a jet from a top decay (top boosted → decay collimated)



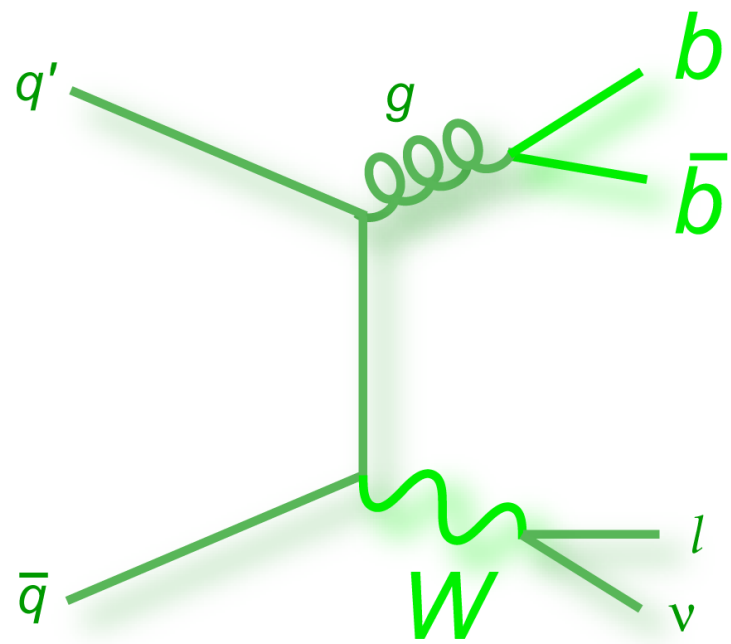
Analysis backgrounds

- ⋆ Background contribution to the analysis :
  - SM $t\bar{t} \rightarrow$ dominant contribution
  - $W+$ jets
  - Multijet
  - + other backgrounds

SM $t\bar{t}$



W + jets

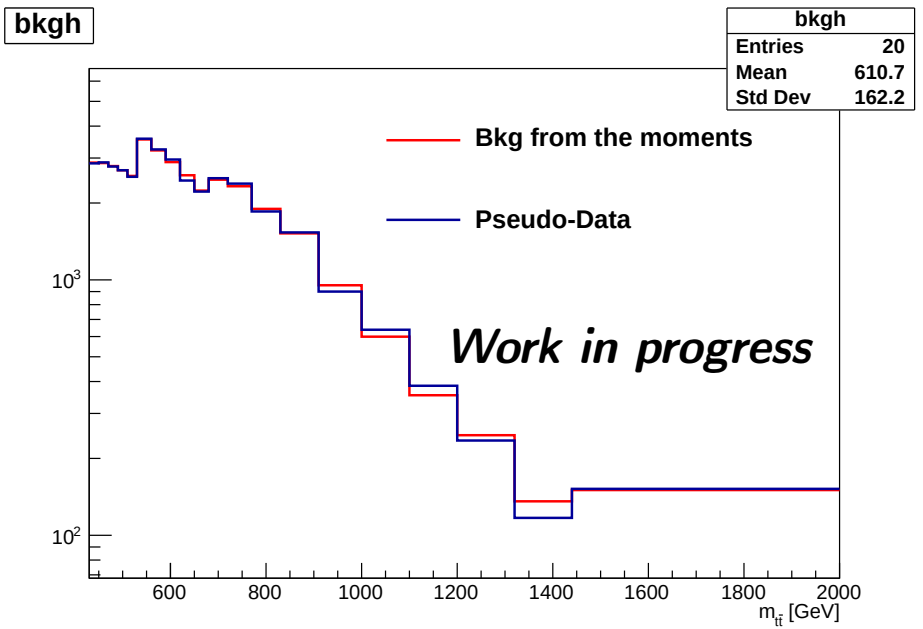


W + jets $\rightarrow$ events with 1 isolated lepton and 1 neutrino from W boson decay

BumpHunter

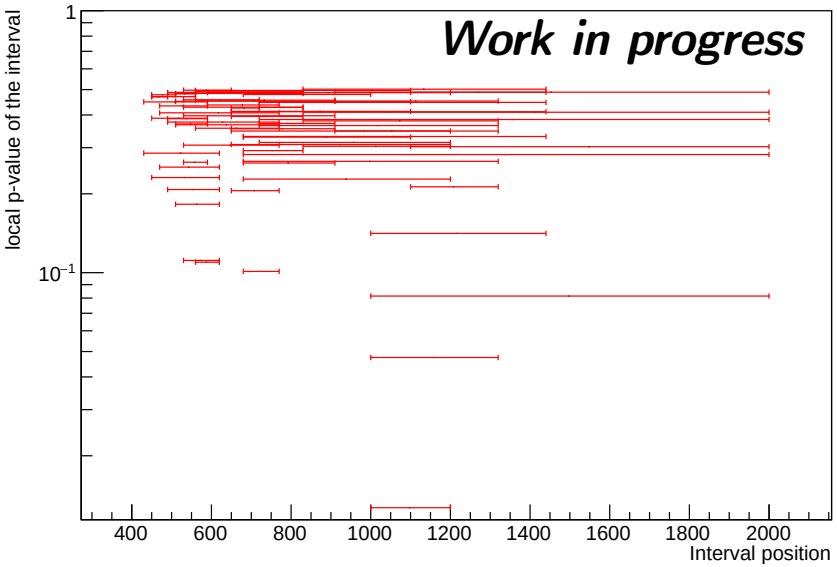
- ⋆ Software to search for excess/deficit in a spectrum :
  - no signal assumption
  - removes the *Look eslwhere Effect*

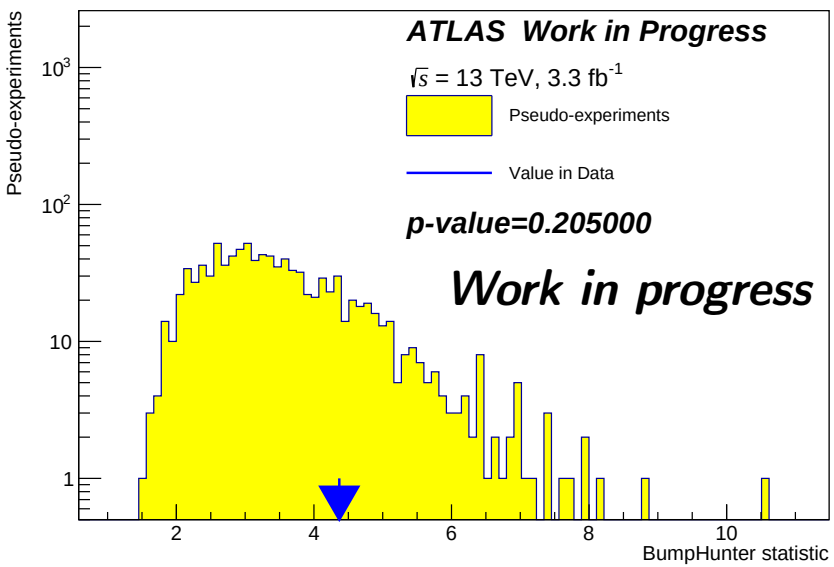
- ⋆ Consider all possible windows (position and width)
- ⋆ Count data d_i and bkg b_i in all windows
- ⋆ Compute the probability that the bkg has fluctuated
  $\rightarrow$ local p-value



BumpHunter

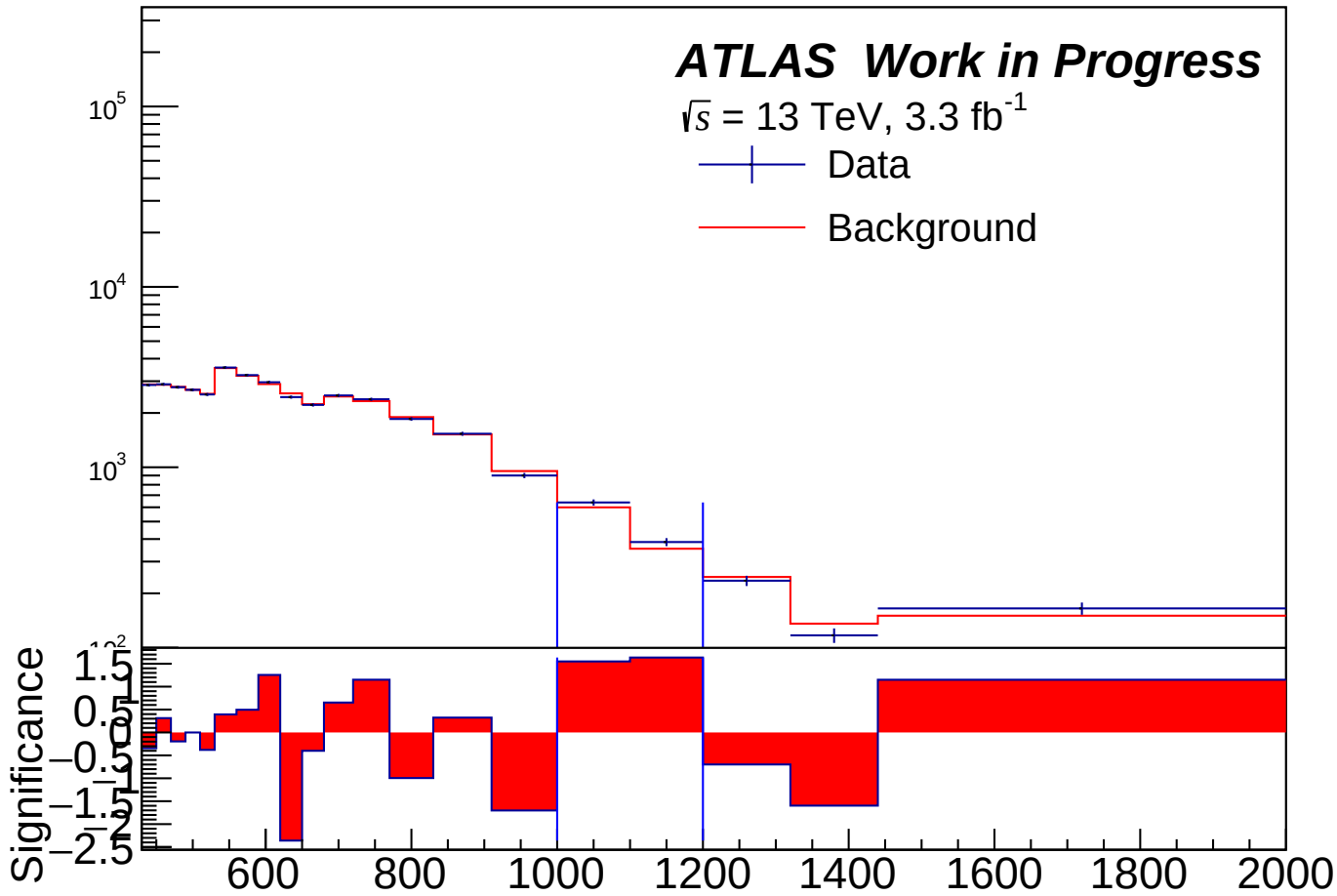
- ⋆ Distribution of local p-value
- ⋆ Lowest local p-value is used to compute BH test statistic : $t = -\log(\text{p-val}^{\text{min}})$
- ⋆ Observed test statistic compatible with bkg hypothesis ?





- ⋆ Pseudo-experiments (PE) generated by MC simulation $\rightarrow$ $\text{p-val}^{\text{min}}$ for each PE
- ⋆ Compute a global p-value : $\text{p-value}^{\text{global}}$ = fraction of $t_{PE} \geq t_{obs}$

BumpHunter result



Return the intervall with the most significant excess/deficit with a global p-value

Upper limits from TRex Fitter

Work in progress

Upper limit (pb)
rmu3 channel only

		SM only	SM+Z' (750 GeV, σ=1.88 pb)
MC	expected	$1.88^{2.62}_{1.35}$	$1.88^{2.62}_{1.35}$
	observed	1.68	3.34
New FD	expected	$1.87^{2.62}_{1.35}$	$1.91^{2.67}_{1.38}$
	observed	1.80	1.86
Old FD	expected	$1.87^{2.61}_{1.35}$	$1.87^{2.61}_{1.35}$
	observed	1.94	3.76

The new FD can't see the signal
The old FD has similar sensitivity to MC based bkg